PuntoSicuro

ARTICOLO DI PUNTOSICURO

Anno 28 - numero 6197 di Mercoledì 18 novembre 2026

Deepfake, una nuova norma ETSI per prevenire i rischi dell'intelligenza artificiale

La specifica tecnica ETSI TS 104 275 introduce misure concrete per prevenire la creazione e la diffusione di deepfake: valutazione dei rischi, controllo umano, protezione dei contenuti e monitoraggio continuo dei sistemi di IA.

Alla fine di ottobre 2026 ETSI ha pubblicato una specifica tecnica, che mira a ridurre i rischi legati al fatto che i sistemi di intelligenza artificiale generativi possono produrre immagini contraffatte "deepfakes".

Questa norma è stata sviluppata dalla commissione tecnica, che è specificamente incaricata di rendere più sicura l' intelligenza artificiale, ed accresce la responsabilizzazione di coloro che sviluppano applicativi, offrendo delle misure pratiche di salvaguardia, che mostrano come sia possibile mitigare i rischi legati ad immagini ed audio contraffatti.

Sino ad oggi gli sforzi per combattere le immagini contraffatte si erano concentrate inizialmente sulla individuazione della manipolazione del contenuto, dopo che è stato creato. Questa nuova norma assume un atteggiamento proattivo, facendo in modo che la prevenzione della contraffazione sia una sfida di sicurezza, che deve essere fronteggiata attraverso l'intero ciclo di vita dell'applicativo AI.

Questa norma è stata sviluppata, facendo riferimento ai requisiti di sicurezza informatica, che già erano stati individuati nella norma ETSI EN 304223, applicabili a sistemi e modelli di **AI**.

Pubblicità

In particolare, questa norma sottolinea le seguenti misure di prevenzione:

- **effettuare una valutazione di rischio prima dello sviluppo dell'applicativo:** è così possibile identificare possibili usi non corretti degli applicativi generativi e attuare delle salvaguardie, prima che questi applicativi vengano diffusi,
- **controllo umano:** vengono indicati dei processi che permettono di introdurre una sorveglianza umana, in modo da garantire che l'organizzazione mantenga il controllo sul comportamento di questi applicativi,
- **protezioni contro usi non appropriati**: la norma dà delle indicazioni su come sia possibile filtrare il contenuto, etichettare il prodotto generato dall'applicativo e mettere in guardia gli utenti, quando appropriato,
- **monitoraggio continuo**: la norma indica come tenere sotto controllo nuovi rischi, indagando su eventuali incidenti verificatisi e aggiornando le misure di salvaguardia, mano a mano che nuovi rischi vengono identificati,
- **valutazioni basate su elementi oggettivi**: la norma indica come sia appropriato mantenere una documentazione ed una serie di registri, che dimostrino come siano stati attuati appropriati controlli, secondo le indicazioni della norma.

Oltre a queste misure di salvaguardia, la norma indica anche degli scenari di prova e dei criteri di tipo si/no. Queste misure pratiche consentono all'organizzazione di dimostrare che hanno attuato un efficiente ed efficace programma di contrasto delle contraffazioni, trasformando una declaratoria in una affermazione documentata.

ETSI TS 104 275 V1.1.1 (2026-09) - Securing Artificial Intelligence (SAI); Application of Baseline Cyber Security Requirements for AI Models and Systems principles to Generative AI; Preventing Harm from Deepfakes - Sicurezza dell'Intelligenza Artificiale (SAI); Applicazione dei principi dei Requisiti di Sicurezza Informatica di base per i Modelli e i Sistemi di IA all'Intelligenza Artificiale Generativa; Prevenzione dei danni derivanti dai Deepfake

Adalberto Biasiotti

Licenza Creative Commons
I contenuti presenti sul sito PuntoSicuro non possono essere utilizzati al fine di addestrare sistemi di intelligenza artificiale.

www.puntosicuro.it